2016. Вып. 2-3

УДК 579.519.6 doi: 10.18101/2306-2363-2016-2-3-10-17

© D. Batkhishig, P. Enkhbayar

HELICAL PARAMETERS OF REGULAR π -HELICES IN PROTEINS

The α -helix, 3_{10} -helix, π -helix and ω -helix have been observed in protein structures; they account for 32% of residues, 4%, 0.3% and 0.2%, respectively. However these percentages depend on resolution of solved structures and method for assignment of secondary structures. Culled data set, containing 2901 protein chains with less than 25% sequence identity and ≤ 1.6 Å resolution (*R*-value \leq 0.25), was used in this analysis. Secondary structure assignments are performed by DSSP, STRIDE, and SECSTR for π -helices.

HELFIT program determines the helical parameters-pitch, residues per turn, radius, and handedness and $p = rmsd/(N-1)^{1/2}$ for π -helices, where RMSD is the root mean squares deviation from the best fit helix and N is helix length. p-value, estimates helical regularity and all regular π -helices with $p \leq 0.10$ Å were identified. Helical parameters of π -helices are compared with the helical parameters of canonical π -helices and other types of protein helices.

Keywords: 3_{10} -helix, α -helix, π -helix, helical parameters, regular helix, protein structures, protein chains.

Д. Батхишиг, П. Энхбаяр

СПИРАЛЬНЫЕ ПАРАМЕТРОВ РЕГУЛЯРНЫХ *п*-СПИРАЛЕЙ В БЕЛКАХ

а-Спираль, 3_{10} -спираль, π -спираль и ω -спираль были отмечены в белковых структурах. На них приходится 32% от остатка, 4%, 0.3% и 0.2%, соответственно. Однако эти проценты зависят по разрешения разрешающих структур и способа для присвоения вторичной структуры. Отобранный набор данных, содержащий 2901 белковые цепи с менее, чем 25% последовательной идентичности и разрешением ≤ 1.6 Å (R-значение \leq 0.25), был использован в данном анализе. Вторичные структуры присвоений выполняются DSSP, STRIDE и SECSTR для π -спиралей.

HELFIT программа определяет спиральные параметры-высоту тона, аминокислотный остаток на виток, радиус, и направленность и $p = rmsd/(N-1)^{1/2}$ для π -спиралей, где RMSD является среднеквадратичным отклонением от оптимальной подобранной спирали и N является длиной спирали. p-значение оценивает спиральную регулярность и все регулярные π спирали с p = 0.10Å были определены. Спиральные параметры π -спиралей сопоставлены со спиральными параметрами канонических π -спиралей и другим видам белковых спиралей. Д. Батхишиг, П. Энхбаяр. Спиральные параметров регулярных π-спиралей в белках

Ключевые слова: 3₁₀-спираль, α-спираль, *π*-спираль, спиральные параметры, регулярные спирали, белковые структуры, белковые цепи.

Helix is one of two main types of secondary structures in proteins. Helices are usually designated as i_n based on the number of residues per turn (*i*) and the number of atoms in the ring joined by the backbone hydrogen bond (*n*) [3]. Pauling and Corey first hypothesized the α -helix (3.6₁₃) and the γ -helix (5.1₁₇) structures [15]. Donohue later considered the possibility of other types of helices (2.2, 3₁₀, 4.3₁₄ and 4.4₁₆) [3]. Low and Baybutt also suggested the possibility of the 4.4₁₆helix or π -helix [14]. The main stabilizing factor for helical structures in polypeptides is repeat hydrogen bonds between main chain carbonyl oxygen (C=O) and amide hydrogen (NH) groups with the π -helix characterized by an (*i* \leftarrow *i*+4) pattern, the 3₁₀- and the π -helix by repealing (*i* \leftarrow *i*+3) and (*i* \leftarrow *i*+5) hydrogen bonds, respectively [11].

There are several programs perform assignments of secondary structures based on three-dimensional atomic coordinates of proteins [7, 11]. Among these, DSSP (Kahsch and Sander, 1983) and STRIDE (Frishman and Argos, 1995) are the most widely used. While DSSP identifies helices based on the repeating $(i \leftarrow i+n)$ hydrogen bonds with corresponding to *n* of 3, 4 and 5 for 3_{10} , α - and π -helices, respectively [11], the STRIDE uses both of hydrogen bonds and main chain dihedral angles to define secondary structures. DSSP program identified only 9 unique π -helices from the database of more than 6000 of proteins [19]. Fodje and Karadaghi defined 116 π -helices using their home made program, SECSTR, from the database of 932 high-resolution three-dimensional structures of proteins [8].

These different results can be explained by the following two reasons: 1) Number of solved 3D structures were insufficient by this time 2) Programs to assign of secondary structures are using different methods.

We studied helical parameters of protein helices with HELFIT program and compared with the parameters of canonical π -helices.

Materials and Methods

Composition of Database

The 16 May 2016 culled PDB data set, containing 2969 protein chains with less than 20% sequence identity and 1.6 Å resolution (*R*-value ≤ 0.25), was used in this analysis.

DSSP program

DSSP performs secondary structure assignments by the bonding energy $E \le -0.5$ kcal/mol between C=O of residue *i* and N–H residue *n* (*i* \leftarrow *i*+*n*). The optimal hydrogen bonding energy for mainchain-mainchain N—H···O hydrogen bonds $E_m < -3$ kcal/mol. Hydrogen bond energy depends on both electrostatic interaction N—H···O of atoms and of hydrogen bonds angle θ [11].

They calculate the electrostatic interaction energy between two hydrogen bonding groups by placing partial charges on the C, O $(+q_1, -q_1)$ and N, H $(-q_2, +q_2)$ atoms.

2016. Вып. 2-3

$$E = q_1 q_2 \left(\frac{1}{r} (\text{ON}) + \frac{1}{r} (\text{CH}) - \frac{1}{r} (\text{OH}) + \frac{1}{r} (\text{CN}) \right) \cdot f$$

E interaction potential energy (kcal/mol), $q_1 = 0.42e$ and $q_2 = 0.20e$, *e* is elementary charge ($|e|=1.6\cdot10^{-19}$ C), *r* is distance between two atoms (Å), the dimensional factor f = 332 kcal/ e^2 mol [11].



Fig. 1. Distances used to calculate hydrogen bond energy the Coulomb interaction.

STRIDE program

STRIDE program is designed for protein secondary structure assignment from three-dimensional atomic coordinates based on the combined use of hydrogen bond energy and statistically derived backbone torsional angle information [7]. The hydrogen bond energy $E_{\rm hb}$ is calculated using the empirical energy function derived from the analysis of experimental data on hydrogen bond geometries in crystal structures of amino acids in polypeptide chains;

$$E_{hb} = E_r + E_t + E_{\theta}$$

where, E_r is the distance dependence of the hydrogen bond, and E_t and E_{θ} describe its directional properties.

$$E_{\rm r} = \frac{C}{r^8} + \frac{D}{r^6}$$

where $C = -3E_m r^8_m$ kcal Å⁸/mol, $D = -4E_m r^6_m$ kcal Å⁶/mol, *r* is the distance between the donor and acceptor atoms participating in the hydrogen bond (fig. 1), and E_m and r_m are the optimal hydrogen bond energy and length, respectively [7]. For mainchain-mainchain hydrogen bonds N—H···O, $E_m = -2.8$ kcal/mol and $r_m = 3.0$ Å [11]. The angular terms E_t and E_{θ} have the following forms:

$$E_{\theta} = \cos^{2}\theta \text{ and}$$

$$E_{\theta} = \begin{cases} (0.9 + 0.1\sin 2t_{i})\cos t_{0}, & 0 < t_{i} < 90^{\circ} \\ K_{1}(K_{2} - \cos^{2}t_{i})^{3}\cos t_{0}, & 90^{\circ} < t_{i} < 110^{\circ} \\ 0, & t_{i} > 110^{\circ} \end{cases}$$

where, $K_1 = 0.9/\cos^6 110^\circ$, $K_2 = \cos^2 110^\circ$ and $C_{\alpha}C'NO$ atoms situated in flat because t_i is of r_{OH} in flat (from O atom to H atom drawn radius vector) mapping C' - O of

Д. Батхишиг, П. Энхбаяр. Спиральные параметров регулярных п-спиралей в белках

bond from directional deviation angle, t_0 is between r_{OH} radius vector and flat angle (fig. 2).

SECSTR program

SECSTR is a new addition to the DSSP program that is dedicated to identifying π -helices, which were seldom assigned by older versions of DSSP and STRIDE [8]. The secondary structure assignment methods based on hydrogen bond assignments (DSSP, STRIDE, and SECSTR) produced nearly identical assignments, with more than to 90% [20].



Fig. 2. An illustration of main-chain hydrogen bond geometry as adapted from Boobbyer et al.

HELFIT program

HELFIT is a continuous helical five parameters described with least squares method. Therefore

- 1. the direction vector of the helix axis (a_x, a_y, a_z)
- 2. the radius of the helix r
- 3. radius of the helix and its +/- error r and $\pm \Delta r$
- 4. the pitch of the helix P
- 5. number of residue per turn *n*

This program described the helical parameters with high accuracy for the analysis least four of data points three-dimensional coordinate and these points are coordinate of C_{α} atoms of amino acids of helix type secondary structure in polypeptide chain. Above the helical some parameters specification displacement of the helix axis along per turn or pitch *P*, it also calculates a parameter *p* = rmsd/(*N*-1)^{1/2}, where rmsd is the root mean square distance from the best-fit helix to data points and *N* is the number of data points [6].

Discussion and Results

The α -helix is considered to be the most abundant form of secondary structure, accounting for about 31% of amino acid secondary structure states, while the 3₁₀-helix accounts for about 4% [1, 2]. The π -helix, however, appears to be extremely rare. The rarity of the π -helix has been attributed to its instability due to the following properties: (1) the dihedral angles φ and ψ are unfavorable, lying at the

very edge of an allowed minimum energy region of the Ramachandran plot [17]; (2) the larger radius of the π -helix means that main chain atoms are no longer in van der Waals contact along the helix axis, resulting in a hole too small for a water molecules to fill [14]; (3) a large entropic cost is required to form a helix in which five residues need to be aligned to permit the ($i \leftarrow i+5$) hydrogen bond [16]. A few researchers have, however, found π -helices to be formed during molecular dynamics simulations of peptides [12] with some reports of a transition from α helix to π -helix structure [4]. This suggests that the π -helix is not as unstable as previously believed.

We identified 27, 22 and 340 π -helices from 2901 proteins by DSSP, STRIDE and SECSTR respectively. All π -helices are divided into two types, regular and irregular, with *p*-value: $p \le 0.10$ Å regular and p > 0.10 Å irregular. In order to compare protein π -helices with the canonical π -helices the only parameters of regular π -helices' are used for the further analysis.

Table 1

	Length	P (Å)	n	$\Delta z^{b}(\text{\AA})$	r (Å)	$V_{\rm c}({\rm \AA}^3)^a$	p (Å)
Average	7.71	5.13	4.41	1.16	2.76	27.76	0.07
Standard deviation	0.76	0.10	0.09	0.03	0.04	0.77	0.03
Minimum	7.00	5.01	4.25	1.11	2.68	26.76	0.02
Maximum	9.00	5.30	4.53	1.19	2.81	29.20	0.10
Canonical	-	5.16	4.40	1.15	2.68	25.90	_

Helical parameters of 7 regular π -helices in proteins

^{*a*} Voronoi volume ($V_c = \pi \cdot r^2 \cdot \Delta z$), ^{*b*} Helix rise per residue $\Delta z = P/n$

Helical parameters of π -helices identified by DSSP program

Total 27 (20 irregular π -helices) π -helices are identified from 2901 proteins by DSSP program. Helix length, helical parameters, and *p*-values of π -helices in protein are determined with the HELFIT program (Table 1).

Helix radius of real π -helices is larger than the radius of canonical π -helix, also Voronoi volume of real π -helices is larger than canonical π -helix. The other helix parameters are close to the parameters of canonical p-helix. Average length is 7.71 and length is range of 7-9 residues.

Helical parameters of π -helices identified by STRIDE program

 π -helices (22) are identified the high resolution 3D structures of proteins by DSSP program. Helical parameters, and p-values of π -helices in protein are determined with the HELFIT program and Voronoi volume (V^a) and helix rise per residue (Δz) are calculated of π -helices (Table 2).

Д. Батхишиг, П. Энхбаяр. Спиральные параметров регулярных *п*-спиралей в белках

Table 2

	Length	$P(\text{\AA})$	n	$\Delta z^{b}(\mathbf{A})$	r (Å)	$V_{\rm c}({\rm \AA}^3)^a$	p (Å)
Average	7	5.10	4.46	1.14	2.78	27.65	0.07
Standard deviation	0	0.06	0.06	0.02	0.03	0.36	0.04
Minimum	7	4.99	4.40	1.10	2.74	27.33	0.03
Maximum	7	5.14	4.54	1.17	2.81	28.39	0.14
Canonical	_	5.16	4.40	1.15	2.68	25.90	-

Helical parameters of 6 regular π -helices in proteins

Length of all real π -helices is same and helix radius of real π -helices is larger than the radius of canonical π -helix. Real values of *P*, *r* are close to the parameters of canonical π -helix.

Helical parameters of π -helices identified by SECSTR program

340 π -helices (264 are irregular helices) are identified the high resolution 3D structures of proteins by SECSTR program. Helical parameters of π -helices in protein are determined with the HELFIT program and Voronoi volume (V^{α}) and helix rise per residue (Δz) are calculated of π -helices (Table 3).

Table 3

	Length	$P(\text{\AA})$	п	$\Delta z^{b}(\mathrm{\AA})$	<i>r</i> (Å)	$V_{\rm c}({\rm \AA}^3)^a$	p (Å)
Average	7.42	5.18	4.40	1.18	2.75	28.10	0.08
Standard deviation	1.16	0.12	0.21	0.09	0.06	1.98	0.02
Minimum	5.00	4.82	2.99	1.03	2.58	23.93	0.03
Maximum	12.00	5.51	5.07	1.84	2.88	42.20	0.10
Canonical	_	5.16	4.40	1.15	2.68	25.90	

Helical parameters of 76 regular π -helices in proteins

Real values of *P*, Δz , *r*, and *V*_c, more than those of canonical helix and value of same that of canonical π -helix.

Conclusion

• 2901 3D structures of high resolution protein structures were downloaded from Protein Data Bank (PDB) and there are 389 π -helices. In average, every protein contains 0.13 π -helices.

• All π -helices are divided into two groups, regular and irregular. There are 89 π -helices are regular among the total of 389 π -helices, 4.37%. Helix parameters of all regular π -helices are used for futher analysis.

• Radii of all π -helices are larger than that of canonical π -helices and all the helical parameters are comparable with those of canonical helices.

References

1. Baker E. N., Hubbard R. E. Hydrogen bonding in globular proteins // Prog. Biophys. Mol. Biol. - 1984. - V. 44. - P. 97-179.

2. Barlow D. J., Thornton J. M. Helix geometry in proteins // J. Mol. Biol. - 1988. - V. 201. - P. 601–619.

3. Donohue J. Hydrogen Bonded Helical Configurations of the Polypeptide Chain // Pro. Nat. Acad. Sci. USA. - 1953. - V. 39. - P. 470–478.

4. Duneau J. P., Genest D. and Genest M. Detailed description of an alpha helix-pi bulge transition detected by molecular dynamics simulations of the p185(c-erbB2) V659G transmembrane domain // J. Biomol. Struct. Dyn. - 1996. - V. 13. - P. 753-769.

5. Enkhbayar P., Boldgiv B. and Matsushima N. ω-Helices in Proteins // Protein J. - 2010. - V. 29. - P. 242-249.

6. Enkhbayar P., Damdinsuren S., Osaki M., Matsushima N. HELFIT: Helix fitting by a total least squares method // Comput. Biol. Chem. - 2008. - V. 32. - P. 307–310.

7. Frishman D., Argos P. Knowledge-based protein secondary structure assignment // Proteins. - 1995. - V. 23. - P. 566-579.

8. Fodje M. N., Al-Karadaghi S. Occurrence, conformational features and amino acid propensities for the pi-helix // Protein Eng. - 2002. - V. 15. - P. 353–358.

9. Hobohm U., Scharf M., Schneider R., Sander C. Selection representative set of protein structures // Protein Sci. - 1992. - V. 1. - P. 409–417.

10. IUPAC-IUB, Commission on Biochemical Nomenclature // J. Mol. Biol. - 1970. - V. 52. - P. 1–17.

11. Kabsch W., Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features // Biopolymers. - 1983. - V. 22. - P. 2577–2637.

12. Kovacs H., Mark A. E., Johansson J. and van Gunsteren W. F. The effect of environment on the stability of an integral membrane helix: molecular dynamics simulations of surfactant protein C in chloroform, methanol and water // J. Mol. Biol. - 1995. - V. 247. - P. 808-822.

13. Lee K. H., Benson D. R. and Kuczera K. Transitions from α to π helix observed in molecular dynamics simulations of synthetic peptides // Biochemistry. - 2000. - V. 39. - P. 13737-13747.

14. Low B. W., Baybutt R. B. The pi-helix - a hydrogen bonded configuration of the polypeptide chain // J. Am. Chem. Soc. - 1952. - V. 74. - P. 5806-5814.

Д. Батхишиг, П. Энхбаяр. Спиральные параметров регулярных *п*-спиралей в белках

15. Pauling L., Corey R. B., Branson H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain // Proc. Nat. Acad. Sci. USA. - 1951. - V. 37. - P. 205–211.

16. Perutz M. F. New X-Ray Evidence on the Configuration of Polypeptide Chains // Nature. - 1951. - V. 167. - P. 1053–1054.

17. Ramachandram G. N and Sasisekharan V. Conformation of polypeptides and proteins // Adv. Protein Chem. - 1968. - V. 23. - P. 283-437.

18. Rohl C. A and Doing A. J. Models for the 3(10)-helix/coil, pi-helix/coil, and alpha-helix/3(10)-helix/coil transitions in isolated peptides // Protein Sci. - 1996. - V. 5. - P. 1687-1696.

19. Weaver T. M. The pi-helix translates structure into function // Protein Sci. - 2000. - V. 9. - P. 201–206.

20. Manoj Tyagi, Aure lie Bornot, Bernard Offmann and Alexandre G. de Brevern. Analysis of loop boundaries using different local structure assignment methods // Protein science. - 2009. - V. 18. - P. 1869-1881.

Batkhishig D., Department of Physics, School of Mathematics and Natural Science, Mongolian National University of Education, Laboratory of Bioinformatics and Systems Biology, Department of Information and Computer Science, School of Engineering and Applied Sciences, National University of Mongolia Ulaanbaatar, Mongolia, E-mail: d.batkhishig@msue.edu.mn

Enkhbayar P., Laboratory of Bioinformatics and Systems Biology, Department of Information and Computer Science, School of Engineering and Applied Sciences, National University of Mongolia Ulaanbaatar, Mongolia, E-mail: enkhbayar.p@seas.num.edu.mn

Батхишиг Д., Отделение физики, школа математики и естественных наук, Монгольский национальный педагогический университет, лаборатория биоинформатики и системной биологии, Отделение информационных и компьютерных наук, школа инженерных и прикладных наук, Национальный университет Монголии, Монголия, Улан-Батор, E-mail: d.batkhishig@msue.edu.mn

Энхбаяр П., лаборатория биоинформатики и системной биологии, Отделение информационных и компьютерных наук, школа инженерных и прикладных наук, Национальный университет Монголии, Монголия, Улан-Батор, E-mail: enkhbayar.p@seas.edu.mn